



UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
10/743,547	12/23/2003	Michal Louz-On	0026-0042	5793
44989	7590	09/19/2007		
HARRITY SNYDER, LLP 11350 Random Hills Road SUITE 600 FAIRFAX, VA 22030			EXAMINER TANG, KAREN C	
			ART UNIT 2151	PAPER NUMBER
			MAIL DATE 09/19/2007	DELIVERY MODE PAPER

Please find below and/or attached an Office communication concerning this application or proceeding.

The time period for reply, if any, is set in the attached communication.

Office Action Summary

Application No.

10/743,547

Applicant(s)

LOUZ-ON, MICHAL

Examiner

Karen C. Tang

Art Unit

2151

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) OR THIRTY (30) DAYS, WHICHEVER IS LONGER, FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

- 1) ☒ Responsive to communication(s) filed on 23 December 2003.
- 2a) ☐ This action is **FINAL**. 2b) ☒ This action is non-final.
- 3) ☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

- 4) ☒ Claim(s) 1-29 is/are pending in the application.
- 4a) Of the above claim(s) _____ is/are withdrawn from consideration.
- 5) ☐ Claim(s) _____ is/are allowed.
- 6) ☒ Claim(s) 1-29 is/are rejected.
- 7) ☐ Claim(s) _____ is/are objected to.
- 8) ☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

- 9) ☐ The specification is objected to by the Examiner.
- 10) ☐ The drawing(s) filed on _____ is/are: a) ☐ accepted or b) ☐ objected to by the Examiner.
Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).
Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).
- 11) ☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

- 12) ☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).
- a) ☐ All b) ☐ Some * c) ☐ None of:
- ☐ Certified copies of the priority documents have been received.
 - ☐ Certified copies of the priority documents have been received in Application No. _____.
 - ☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).
- * See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

- ☒ Notice of References Cited (PTO-892)
- ☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)
- ☒ Information Disclosure Statement(s) (PTO/SB/08)
Paper No(s)/Mail Date 9/1/05, 2/7/06
- ☐ Interview Summary (PTO-413)
Paper No(s)/Mail Date. _____
- ☐ Notice of Informal Patent Application
- ☐ Other: _____

1. Claims 1-29 are presented for examination.

DETAILED ACTION

Claim Rejections - 35 USC § 112

The following is a quotation of the second paragraph of 35 U.S.C. 112:

The specification shall conclude with one or more claims particularly pointing out and distinctly claiming the subject matter which the applicant regards as his invention.

Claims 1-9, recites the limitation "the extracted set of URLs" in Claim 1, Lines 4.

There is insufficient antecedent basis for this limitation in the claim. The term should be rewritten to "the set of URLs extracted"

"the extracted set of URLs" in Claim 1, Lines 6. There is insufficient antecedent basis for this limitation in the claim. The term should be rewritten to "the set of URLs extracted"

Claims 15-19, recites the limitation "the extracted URLs" in Claim 15, Lines 6.

There is insufficient antecedent basis for this limitation in the claim. The term should be rewritten to "a extracted URLs"

Claim Rejections - 35 USC § 103

The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

(a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negated by the manner in which the invention was made.

Claims 1-29 are rejected under 35 U.S.C. 103(a) as being unpatentable over Galai et al hereinafter Galai (US 2004/0177015) in view of Najork et al hereinafter Najork (US 6,952,730).

2. Referring to Claims 1-7, 10-13, 15-23, and 25-28, Galai discloses a method for exacting a set of uniform resource locators (URLs) from at least one document (refer to 0003); analyzing the extracted set of URLs to determine those in the set of URLs that contain session identifiers (refer to 0012, 0013, and 0015); generating a clean set of URLs from the extracted set of URLs using the session identifiers (refer to 0019); determining when at least one second URL has already been crawled based, at least in part, on a comparison of the second URL to the clean set of URLs (refer to 0020); wherein the generating a clean set of URLs includes removing the session identifiers to obtain the clean set of URLs (refer to 0019); wherein the at least one document is a web document downloaded from a web site (Crawling function including download the web document from the website from a website.); wherein the session identifiers are determined as including sub-strings from the set of URLs that do not reference content (refer to 0069); wherein the set of URLs are extracted from a web document associated with a web host (refer to 0003); wherein the set of URLs are extracted from multiple web documents associated with a single web host (refer to 0003 and 0004); store clean versions of the extracted URLs in which the session identifiers are removed from the extracted URLs (the clean version must be stored in order to compare with the original version, refer to 0019).

Although Galai disclosed the invention substantially as claimed, Galai is silent regarding “wherein the comparison of the second URL to the clean set of URL is based on a comparison of a fingerprint value calculated for each of the URLs in the clean set of URLs.”

Najork, in an analogous art disclosed “wherein the comparison of the second URL to the clean set of URL is based on a comparison of a fingerprint value calculated for each of the URLs in the clean set of URLs.” (refer to Col 9, Lines 4-17).

Hence, providing features disclosed by Najork, would be desirable for a user to implement to provide an efficient data structures that keep in tracks of downloaded document due to crawling the web pages.

Therefore, at the time of the invention, it would have been obvious to one of ordinary skill in the art to modify the system of Galai by including the features disclosed by Najork.

3. Referring to Claims 8 and 9, Galai further discloses a method of claim 1, downloading content from the second URL when the second URL is determined to not already have been crawled (refer to 0019, 0023, once the determination is made, the webpage is not redundant, then the web page is being retrieved.); storing the extracted set of URLs, including embedded session identifier, for use in later accessing the extracted set of URL (act of crawling provides storing the extracted URL, that includes the session identifier, 0003);

4. Referring to Claims 14, 24 and 29, Galai further discloses a method of claim 13, although Galai disclosed the invention substantially as claimed, Galai is silent regarding “adding a generated session identifier to URLs in the clean set of URLs when the URL are to be used to access a web document.”

Najork, in an analogous art disclosed, “adding a generated session identifier to URLs in the clean set of URLs when the URL are to be used to access a web document.” (refer to Col 6, Lines 55-67).

Hence, providing features disclosed by Najork, would be desirable for a user to implement to provide an efficient data structures that keep in tracks of downloaded document due to crawling the web pages.

Therefore, at the time of the invention, it would have been obvious to one of ordinary skill in the art to modify the system of Galai by including the features disclosed by Najork.

Conclusion

5. **Examiner’s Notes:** Examiner has cited particular columns and line numbers in the references applied to the claims above for the convenience of the applicant. Although the specified citations are representative of the teachings of the art and are applied to specific limitations within the individual claim, other passages and figures may apply as well. It is respectfully requested from the applicant in preparing responses, to fully consider the references in entirety as potentially teaching all or part of the claimed invention, as well as the context of the passage as taught by the prior art or disclosed by

the Examiner. In the case of amending the claimed invention, Applicant is respectfully requested to indicate the portion(s) of the specification which dictate(s) the structure relied on for proper interpretation and also to verify and ascertain the metes and bounds of the claimed invention.

A shortened statutory period for reply to this Office action is set to expire THREE MONTHS from the mailing date of this action.

Any inquiry concerning this communication or earlier communications from the examiner should be directed to Karen C. Tang whose telephone number is (571)272-3116. The examiner can normally be reached on M-F 7 - 3.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Valencia Martin-Wallace can be reached on (571)272-3440. The fax phone number for the organization where this application or proceeding is assigned is 571-273-8300.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see <http://pair-direct.uspto.gov>. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free). If you would like assistance from a USPTO Customer Service Representative or access to the automated information system, call 800-786-9199 (IN USA OR CANADA) or 571-272-1000.

V. Martin-Wallace
SPE AU 2151